
Functional human variation

13.03.2013

Agata Wesolowska-Andersen

We are all different!

The human genome is $\sim 3 \times 10^9$ base pairs of DNA
 \Rightarrow no two humans ever have been or will be genetically identical

Between any two humans, the amount of genetic variation is $\sim 0.1\%$

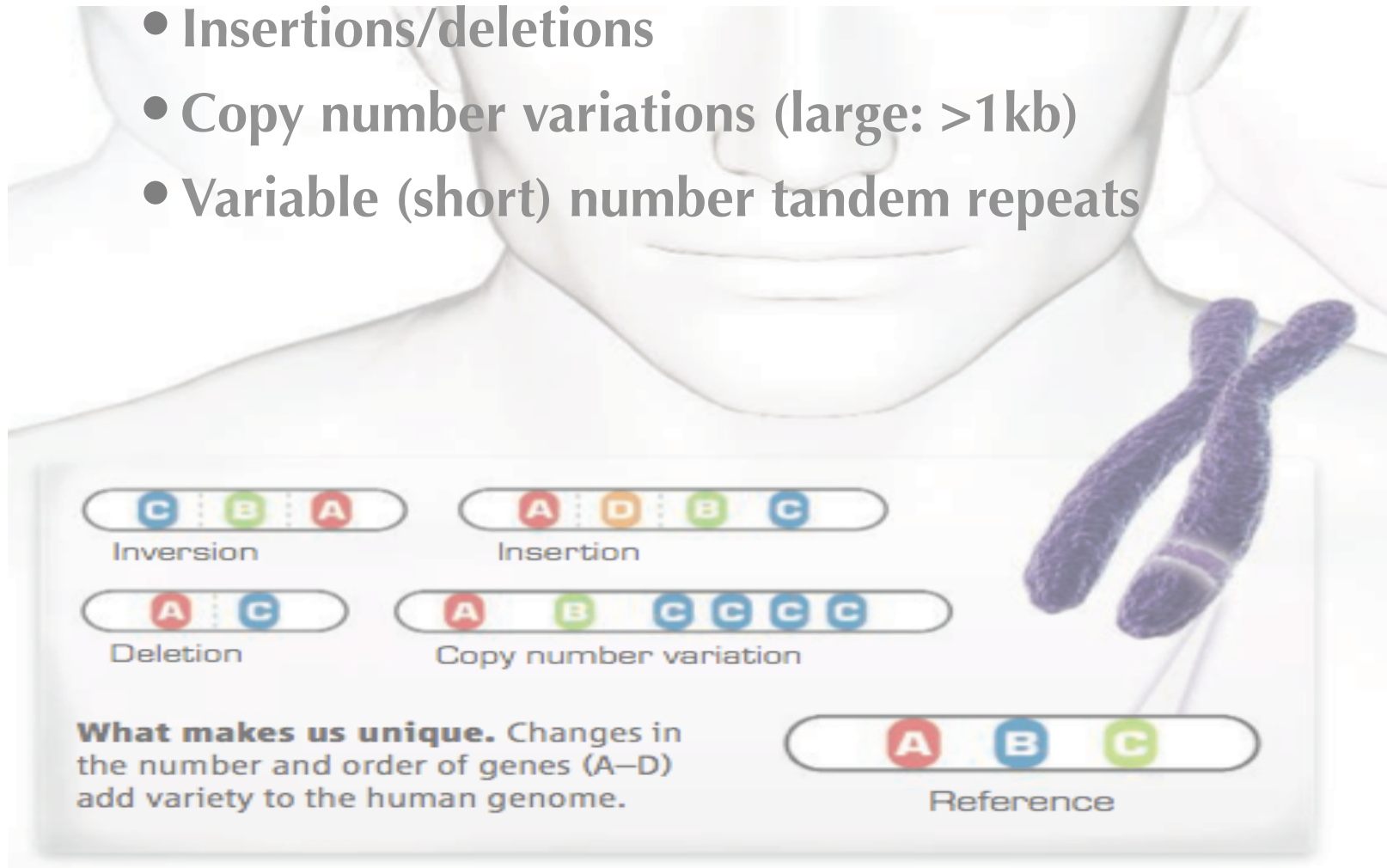
\sim one base pair out of every 1,000 will be different between any two individuals

Any two (diploid) people have about 6×10^6 base pairs that are different!



Genetic variation

- Single nucleotide polymorphisms (SNP)
- Insertions/deletions
- Copy number variations (large: >1kb)
- Variable (short) number tandem repeats



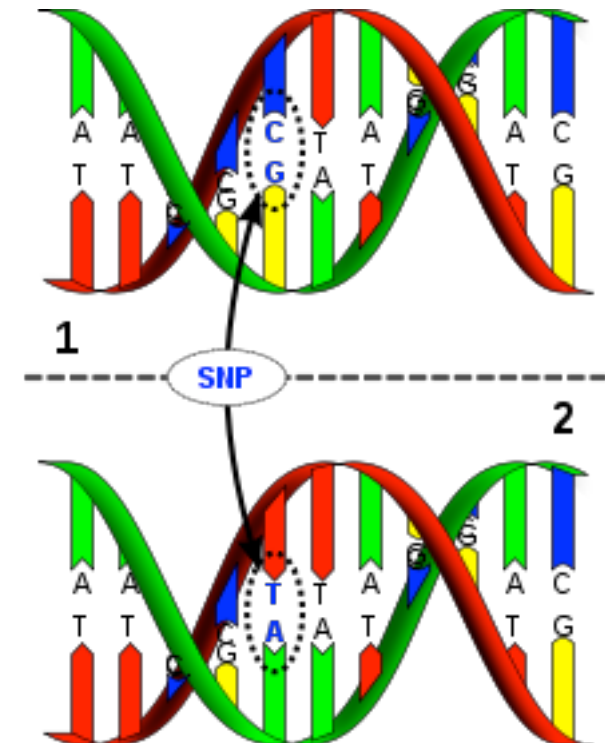
SNPs

SNPs (single nucleotide polymorphism) are the most common type of genetic variation between people

SNP is a one nucleotide difference in DNA sequence occurring in at least 1 % of the population

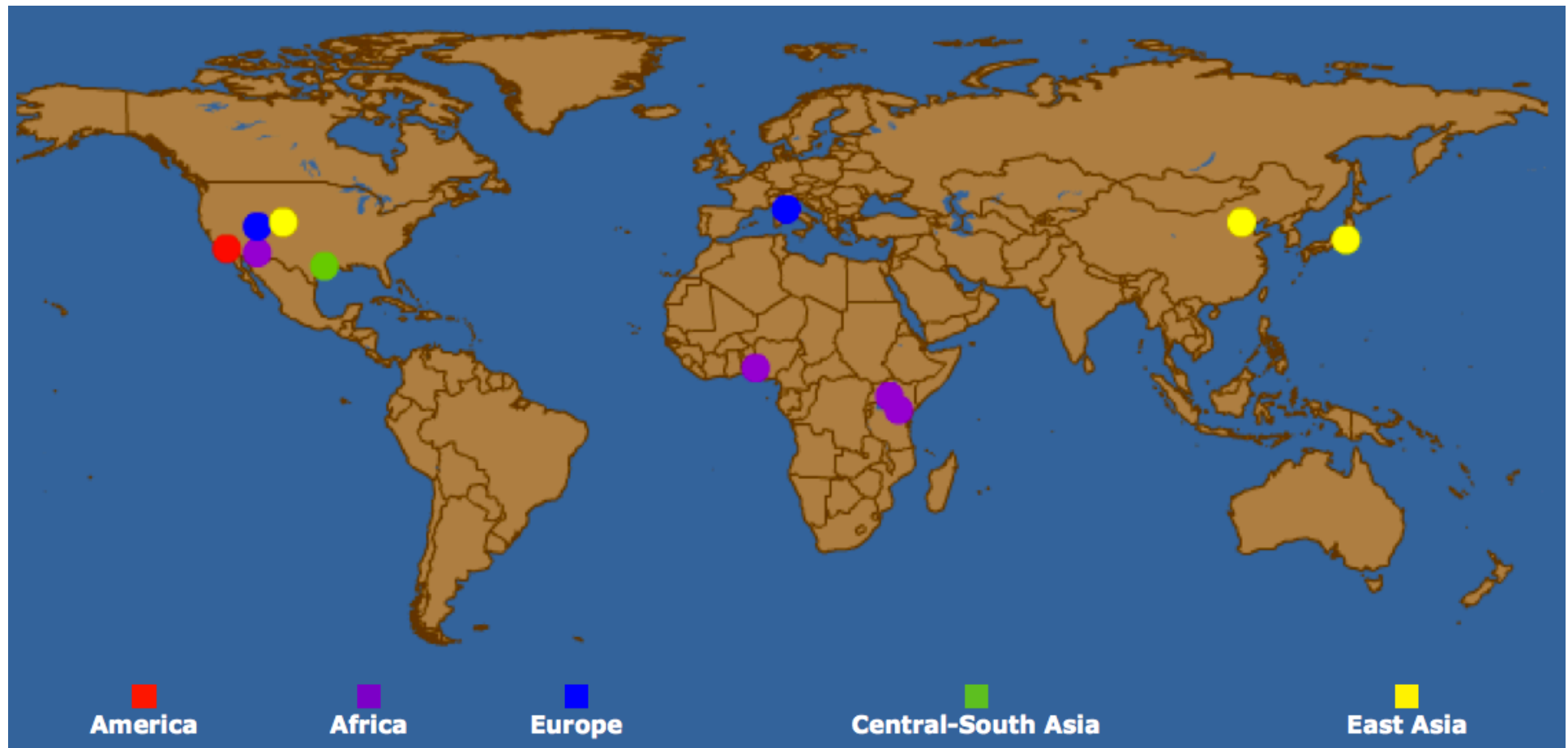
There is ~ 10 - 30 million SNPs in the human genome

SNPs occur every 100-300 bases along the 3-billion-base human genome



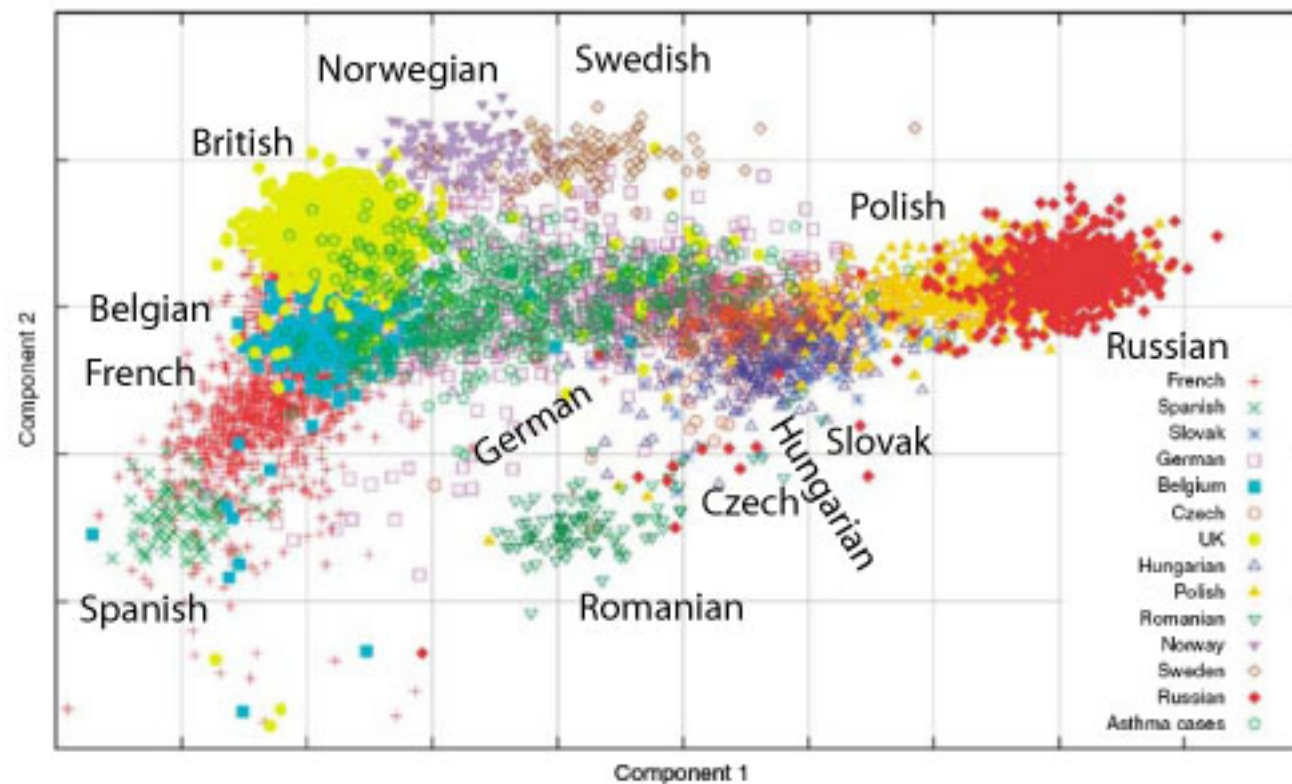
HapMap project

- haplotype map (HapMap) of the human genome
- describes common patterns of human genetic variation



European genetic map

6,000 individuals from 13 populations
~ 300,000 SNPs
PCA



European Journal of Human Genetics (2008) 16, 1413–1429; doi:10.1038/ejhg.2008.210

SNP information

- rs number
- location (chr:position)

Reference SNP(refSNP) Cluster Report: rs17822931 ** With probable-pathogenic... "http://www.ncbi.nlm.nih.gov/sites/varvu?gene=85320&rs=17822931" [detail] **

http://www.ncbi.nlm.nih.gov/SNP/snp_ref.cgi?rs=17822931

NCBI dbSNP Short Genetic Variations

PubMed Nucleotide Protein Genome Structure PopSet Taxonomy OMIM Books SNP

Search Entrez SNP for Go

Have a question about dbSNP? Try searching the SNP FAQ Archive! Go

GENERAL

RSS Feed

Contact Us

Site Map

dbSNP Homepage

Announcements

dbSNP Summary

FTP Download

HUMAN VARIATION

SNP SUBMISSION

DOCUMENTATION

SEARCH

RELATED SITES

Reference SNP(refSNP) Cluster Report: rs17822931 ** With probable-pathogenic allele [detail] **

RefSNP

Organism: human (Homo sapiens)

Molecule Type: Genomic

Created/Updated in build: 123/135

Map to Genome Build: 37.3

Validation Status:

Citation: PubMed

Allele

Variation Class: SNV: single nucleotide variation

RefSNP Alleles: C/T

Allele Origin: G: Germline A: Germline

Ancestral Allele: C

Clinical Source: VarView OMIM

Clinical Significance: With probable-pathogenic allele [detail]

MAF/MinorAlleleCount: T=0.310/679

MAF Source: 1000 Genomes

HGVS Names

NC_000016.9:g.48258198C>T

NG_011522.1:g.15891G>A

NM_032583.3:c.538G>A

NM_033151.3:c.538G>A

NM_145186.2:c.538G>A

NP_115972.2:p.Gly180Arg

NP_149163.2:p.Gly180Arg

NP_660187.1:p.Gly180Arg

Links, Linkout

SNP Details are organized in the following sections:

GeneView Map Submission FaSta Resource Diversity Validation

Integrated Maps (Hint: click on 'Chr Pos' or 'Contig Pos' column value to see variation in NCBI sequence viewer)

Assembly	Genome Build	Chr	Chr Pos	Contig	Contig Pos	SNP to Chr	Contig allele	Contig to Chr	Neighbor SNP	Map Method
GRCh37.p5	37.3	16	48258198	NT_010496.15	1872397	+	C	+	view	blast
reference	36.3	16	48815899	NT_010496.15	1872397	+	C	+	view	blast
Celera	36.3	16	32765323	NW_926462.1	1830122	+	C	+	view	blast
HuRef	37.3	16	34148352	NW_001838286.2	3012385	+	G	-	view	blast
HuRef	36.3	16	34148352	NW_001838286.2	3012385	+	G	-	view	blast

GeneView

GeneView via analysis of contig annotation: **ABCC11** ATP-binding cassette, sub-family C (CFTR/MRP), member 11

View more variation on this gene (click to hide).

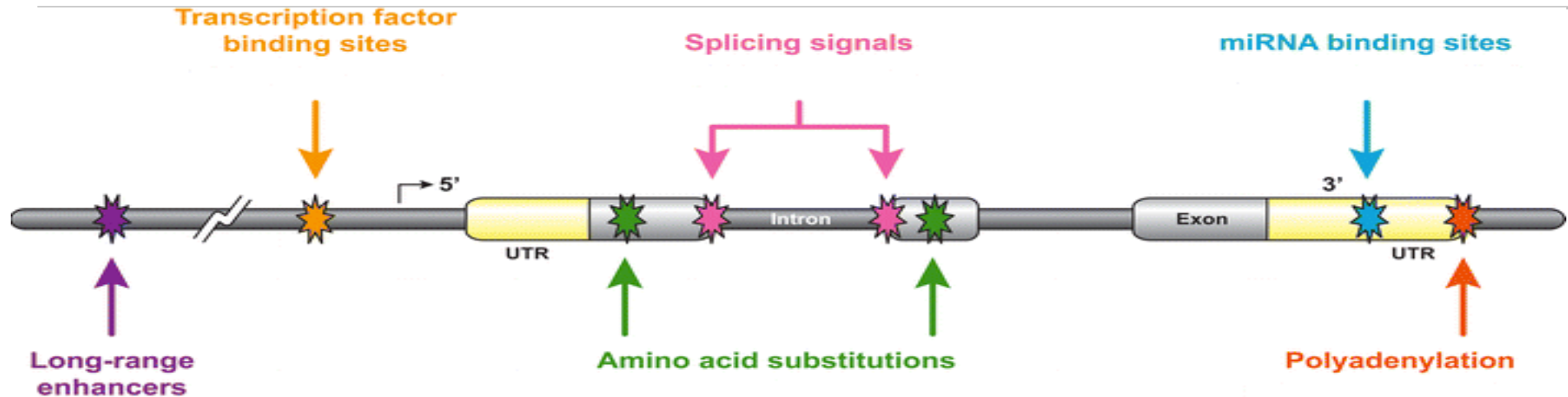
Clinical Source: ☐ in gene region ☒ cSNP ☐ has frequency ☐ double hit Go

Assembly	SNP to Chr	Chr	Chr position	Contig	Contig position	Allele
GRCh37.p5	+	16	48258198	NT_010496.15	1872397	C

RefSeqGene	Gene (ID)	SNP to RefSeqGene	Position	Allele
NG_011522.1	ABCC11 (85320)	-	15891	G

Gene Model(s)

SNP consequence



Coding SNPs:

- non-synonymous coding (missense)
 - change of an amino acid
- synonymous coding (silent mutation) – no change in protein sequence
- stop codon (nonsense) - premature stop codon
- frame-shift coding – insertion/deletion resulting in frame shift

Non-coding SNPs:

Mostly non-functional, but may affect:

- regulatory region
- splice site
- transcription factor binding
- messenger RNA degradation
- sequence of non-coding RNA

SNPs and mutations

Mutations are differences in DNA sequence in an individual that are rare, and may be unique to the individual (or their family line).

Polymorphisms are differences in DNA sequence that are found in many individuals, at a specified frequency (usually 1% or greater of a population)

Polymorphisms start as mutations, but if they become “fixed” in the population, and achieve sufficient frequency, they become polymorphisms.

SNP and genotype

Human are diploid (we have 2 homologous copies of each chromosome)

At each SNP there are 3 possible genotypes:

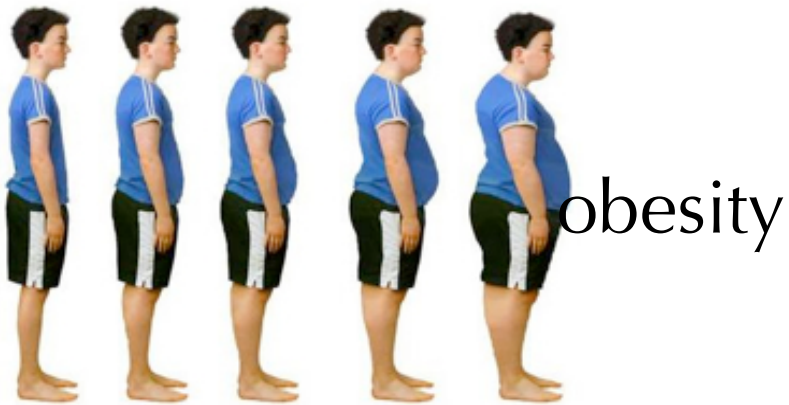
- homozygous reference
- heterozygous
- homozygous non-reference

Homozygous = 2 identical alleles at given locus, eg. AA

Heterozygous = 2 different alleles at given locus, eg. AC

Phenotype

Phenotype = organism's observable characteristic or trait



obesity

eye color



height



An organism's genotype is a major influencing factor in the development of its phenotype

Genotype => phenotype

Genotype

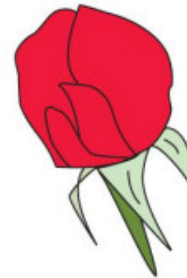
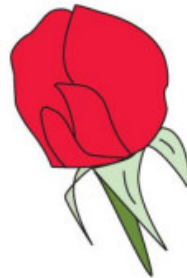
AA

Aa

aa

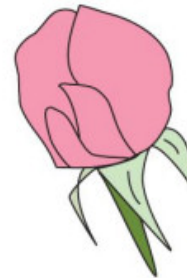
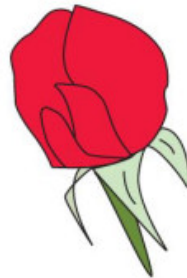
Phenotype

***A*: dominant/*a*:
 recessive**



Phenotype

***A* and *a* codominant**





Navigation

- SNPedia
- Promethease
- FAQ
- Blog
- Current events
- Recent changes

Page [Discussion](#)

Earwax

[rs17822931](#) determines wet vs dry earwax [[PMID 16444273](#)]

This can also be used to distinguish asian ancestry.

[NCBI coffeebreak](#) introduction

Category: Is a medical condition


Genotype	Effect
rs17822931(C;C)	wet earwax
rs17822931(C;T)	wet earwax
rs17822931(T;T)	dry earwax

Direct-to-customer genetic testing

- 23andme
- deCODEme
- Navigenics

~ 1 million SNPs

Start filling in the gaps with your DNA



"Because I had given my doctor information from 23andme, he got to a diagnosis much faster. 23andme saved my life." Kirk C.


\$99*

Our new low price for all!
Was \$199

Order Now »

*Requires a 1-year commitment to the [Personal Genome Service®](#) at \$9/mo. [Order for \\$399](#) without commitment.

1 Get Your Kit 2 Provide Saliva 3 Learn About Yourself 4 Get Monthly DNA Discoveries



Type 2 Diabetes

AS LOW AS 8 % AS HIGH AS 52 % What's your genetic risk? [see more](#)

Gain insight into your traits, from baldness to muscle performance. Discover risk factors for 97 diseases. Know your predicted response to drugs, from blood thinners to coffee. And uncover your ancestral origins. [start tour »](#)

Overview Discover Health & Ancestry Keep Your Doctor Informed Participate In Research

Association studies

Affected



Unaffected



Test whether allele frequency of a SNP is significantly different between the two groups

GWAS results = p-values

Manhattan plot displays all SNPs on x-axis (ordered by genomic location), and $-\log_{10}$ of their p-values on y-axis

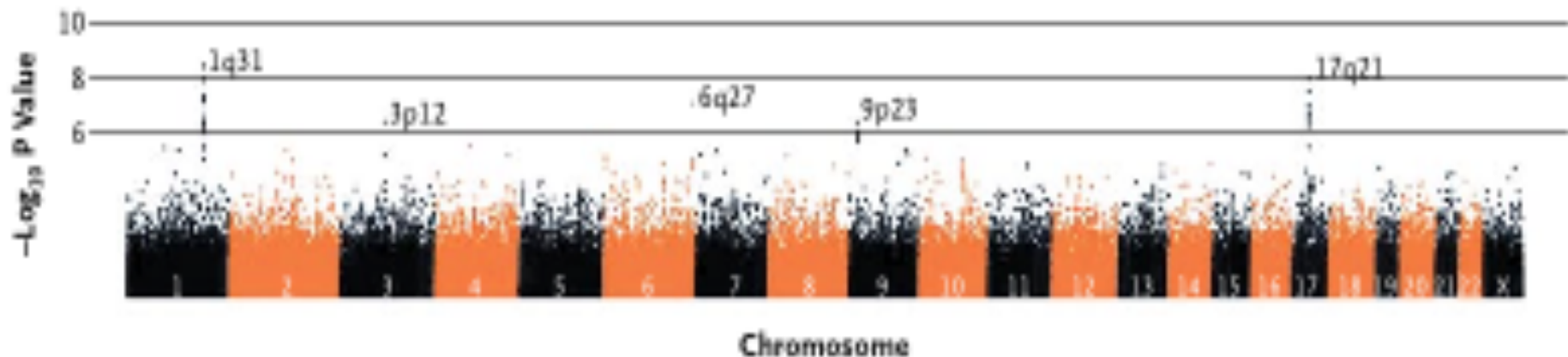
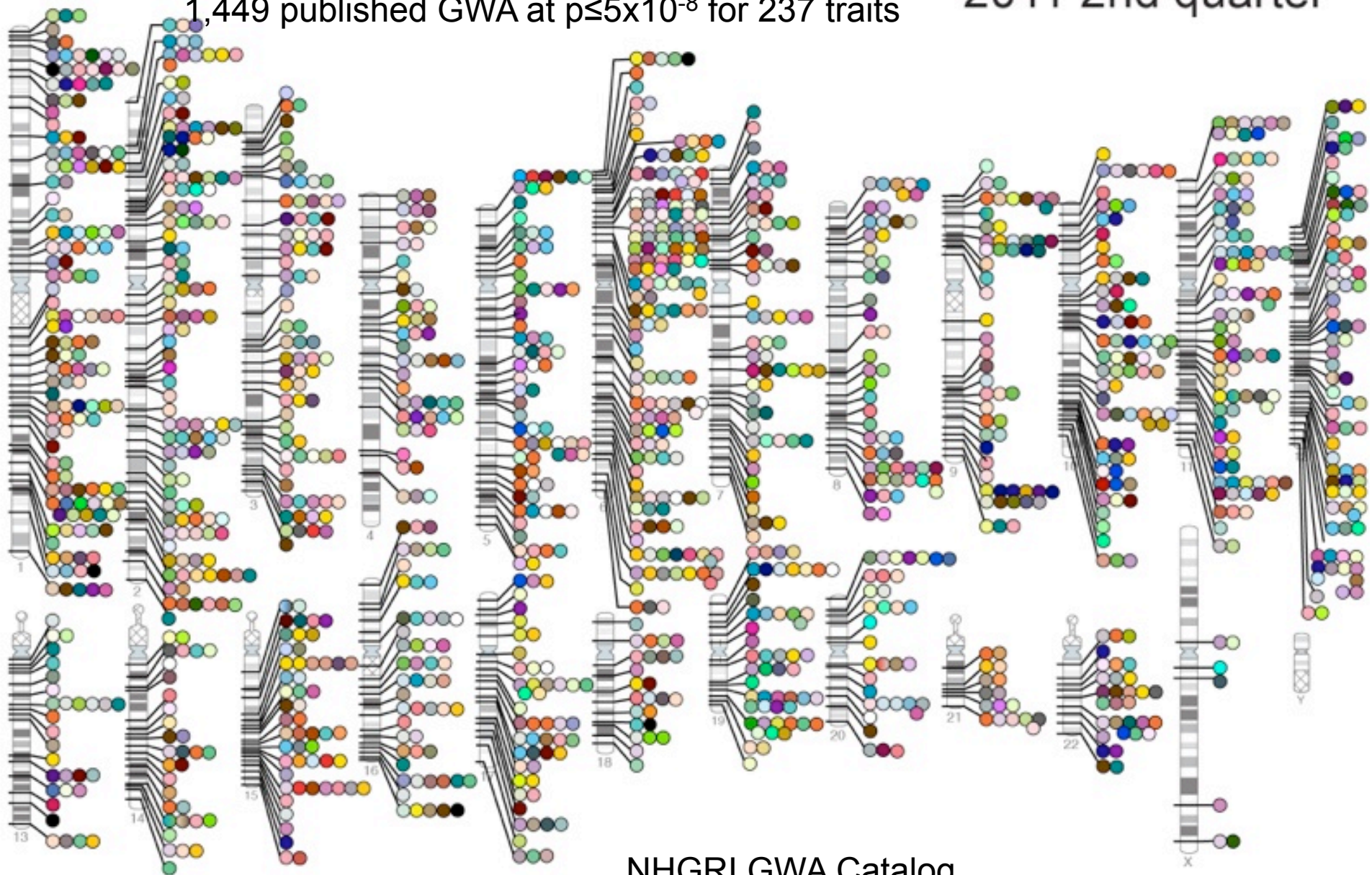


Figure 2. Manhattan Plot of the Results from the Combined Subjects of European Ancestry Who Had Asthma.

The $-\log_{10} P$ values are plotted against the physical distance. Only the two loci at chromosome 1q31 and 17q21 were significantly associated with asthma after Bonferroni correction. Individual chromosome labels are indicated in white within the Manhattan plot.

Published Genome-Wide Associations through 06/2011,
1,449 published GWA at $p \leq 5 \times 10^{-8}$ for 237 traits

2011 2nd quarter



NHGRI GWA Catalog

www.genome.gov/GWAStudies

- Abdominal aortic aneurysm
- Acute lymphoblastic leukemia
- Adhesion molecules
- Adiponectin levels
- Age-related macular degeneration
- AIDS progression
- Alcohol dependence
- Alopecia areata
- Alzheimer disease
- Amyloid A levels
- Amyotrophic lateral sclerosis
- Angiotensin-converting enzyme activity
- Ankylosing spondylitis
- Arterial stiffness
- Asparagus anosmia
- Asthma
- Atherosclerosis in HIV
- Atrial fibrillation
- Attention deficit hyperactivity disorder
- Autism
- Basal cell cancer
- Behçet's disease
- Bipolar disorder
- Biliary atresia
- Bilirubin
- Bitter taste response
- Birth weight
- Bladder cancer
- Bleomycin sensitivity
- Blond or brown hair
- Blood pressure
- Blue or green eyes
- BMI, waist circumference
- Bone density
- Breast cancer
- C-reactive protein
- Calcium levels
- Cardiac structure/function
- Cardiovascular risk factors
- Carnitine levels
- Carotenoid/tocopherol levels
- Celiac disease
- Celiac disease and rheumatoid arthritis
- Cerebral atrophy measures
- Chronic lymphocytic leukemia
- Chronic myeloid leukemia
- Cleft lip/palate

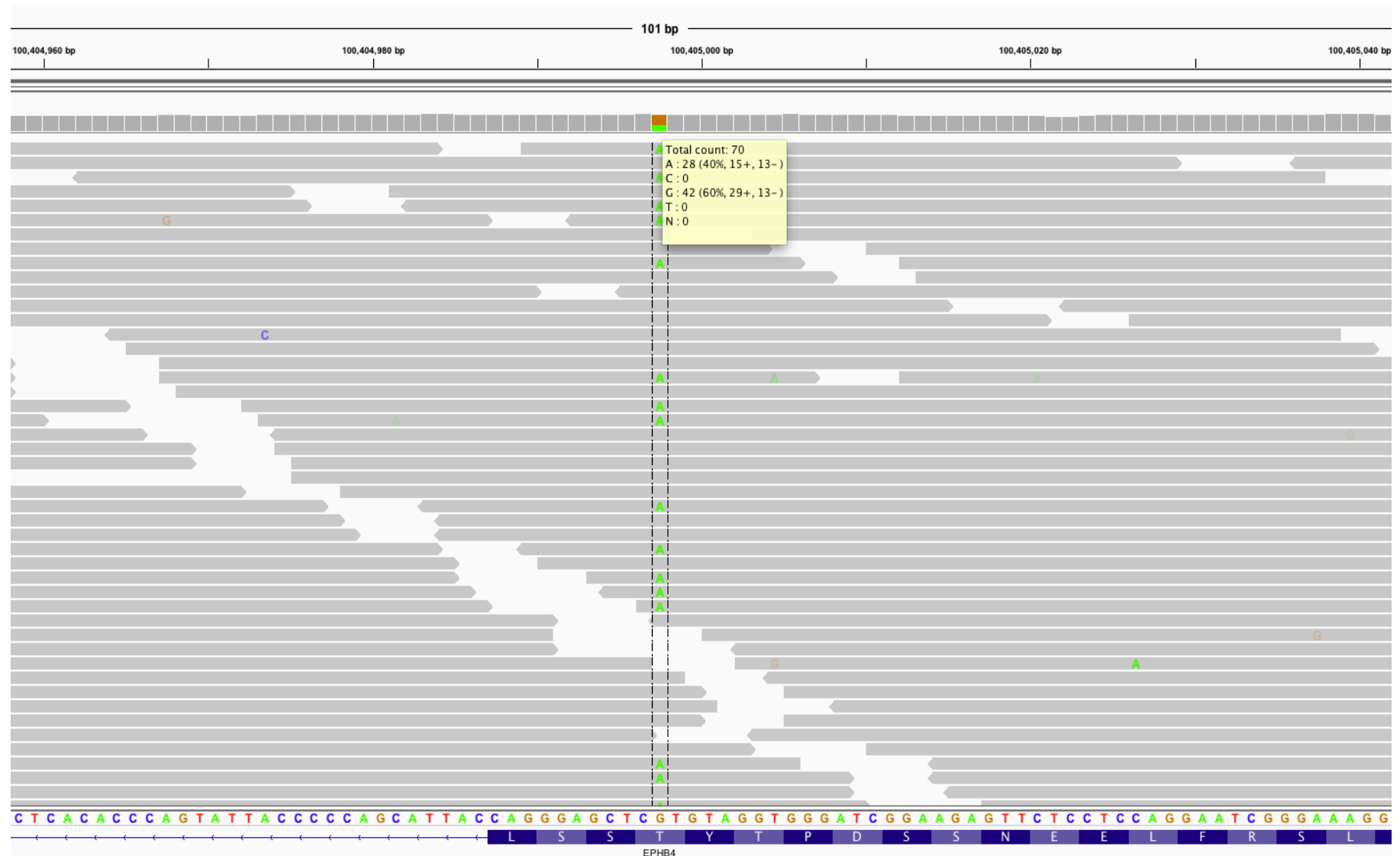
- Coffee consumption
- Cognitive function
- Conduct disorder
- Colorectal cancer
- Corneal thickness
- Coronary disease
- Creutzfeldt-Jakob disease
- Crohn's disease
- Crohn's disease and celiac disease
- Cutaneous nevi
- Cystic fibrosis severity
- Dermatitis
- DHEA-s levels
- Diabetic retinopathy
- Dilated cardiomyopathy
- Drug-induced liver injury
- Drug-induced liver injury (acetaminophen overdose)
- Endometrial cancer
- Endometriosis
- Eosinophil count
- Eosinophilic esophagitis
- Erectile dysfunction and prostate cancer treatment
- Erythrocyte parameters
- Esophageal cancer
- Essential tremor
- Exfoliation glaucoma
- Eye color traits
- F cell distribution
- Fibrinogen levels
- Folate pathway vitamins
- Follicular lymphoma
- Fuch's corneal dystrophy
- Freckles and burning
- Gallstones
- Gastric cancer
- Glioma
- Glycemic traits
- Hair color
- Hair morphology
- Handedness in dyslexia
- HDL cholesterol
- Heart failure
- Heart rate
- Height
- Hemostasis parameters
- Hepatic steatosis
- Hepatitis

- Hepatocellular carcinoma
- Hirschsprung's disease
- HIV-1 control
- Hodgkin's lymphoma
- Homocysteine levels
- Hypospadias
- Idiopathic pulmonary fibrosis
- IFN-related cytopenia
- IgA levels
- IgE levels
- Inflammatory bowel disease
- Insulin-like growth factors
- Intracranial aneurysm
- Iris color
- Iron status markers
- Ischemic stroke
- Juvenile idiopathic arthritis
- Keloid
- Kidney stones
- LDL cholesterol
- Leprosy
- Leptin receptor levels
- Liver enzymes
- Longevity
- LP (a) levels
- LpPLA(2) activity and mass
- Lung cancer
- Magnesium levels
- Major mood disorders
- Malaria
- Male pattern baldness
- Mammographic density
- Matrix metalloproteinase levels
- MCP-1
- Melanoma
- Menarche & menopause
- Meningococcal disease
- Metabolic syndrome
- Migraine
- Moyamoya disease
- Multiple sclerosis
- Myeloproliferative neoplasms
- Myopia (pathological)
- N-glycan levels
- Narcolepsy
- Nasopharyngeal cancer
- Natriuretic peptide levels

- Neuroblastoma
- Nicotine dependence
- Obesity
- Open angle glaucoma
- Open personality
- Optic disc parameters
- Osteoarthritis
- Osteoporosis
- Otosclerosis
- Other metabolic traits
- Ovarian cancer
- Pancreatic cancer
- Pain
- Page's disease
- Panic disorder
- Parkinson's disease
- Periodontitis
- Peripheral arterial disease
- Personality dimensions
- Phosphatidylcholine levels
- Phosphorus levels
- Photic sneeze
- Phytosterol levels
- Platelet count
- Polycystic ovary syndrome
- Primary biliary cirrhosis
- Primary sclerosing cholangitis
- PR interval
- Progranulin levels
- Progressive supranuclear palsy
- Prostate cancer
- Protein levels
- PSA levels
- Psoriasis
- Psoriatic arthritis
- Pulmonary funct. COPD
- QRS interval
- QT interval
- Quantitative traits
- Recombination rate
- Red vs. non-red hair
- Refractive error
- Renal cell carcinoma
- Renal function
- Response to antidepressants
- Response to antipsychotic therapy
- Response to carbamazepine

- Response to clopidogrel therapy
- Response to hepatitis C treatment
- Response to interferon beta therapy
- Response to metformin
- Response to statin therapy
- Restless legs syndrome
- Retinal vascular caliber
- Rheumatoid arthritis
- Ribavirin-induced anemia
- Schizophrenia
- Serum metabolites
- Skin pigmentation
- Smoking behavior
- Speech perception
- Sphingolipid levels
- Statin-induced myopathy
- Stroke
- Sudden cardiac arrest
- Suicide attempts
- Systemic lupus erythematosus
- Systemic sclerosis
- T-tau levels
- Tau AB1-42 levels
- Telomere length
- Testicular germ cell tumor
- Thyroid cancer
- Thyroid volume
- Tooth development
- Total cholesterol
- Triglycerides
- Tuberculosis
- Type 1 diabetes
- Type 2 diabetes
- Ulcerative colitis
- Urate
- Urinary albumin excretion
- Urinary metabolites
- Uterine fibroids
- Venous thromboembolism
- Ventricular conduction
- Vertical cup-disc ratio
- Vitamin B12 levels
- Vitamin D insufficiency
- Vitiligo
- Warfarin dose
- Weight
- White cell count
- White matter hyperintensity
- YKL-40 levels

SNP calling with NGS



VCF file

```
Terminal — ssh — %2
##fileformat=VCFv4.1
##samtoolsVersion=0.1.13 (r926:134)
##INFO=<ID=DP,Number=1,Type=Integer,Description="Raw read depth">
##INFO=<ID=DP4,Number=4,Type=Integer,Description="# high-quality ref-forward bases, ref-reverse, alt-forward and alt-reverse bases">
##INFO=<ID=MQ,Number=1,Type=Integer,Description="Root-mean-square mapping quality of covering reads">
##INFO=<ID=FQ,Number=1,Type=Float,Description="Phred probability of all samples being the same">
##INFO=<ID=AF1,Number=1,Type=Float,Description="Max-likelihood estimate of the site allele frequency of the first ALT allele">
##INFO=<ID=G3,Number=3,Type=Float,Description="ML estimate of genotype frequencies">
##INFO=<ID=HWE,Number=1,Type=Float,Description="Chi^2 based HWE test P-value based on G3">
##INFO=<ID=CI95,Number=2,Type=Float,Description="Equal-tail Bayesian credible interval of the site allele frequency at the 95% level">
##INFO=<ID=PV4,Number=4,Type=Float,Description="P-values for strand bias, baseQ bias, mapQ bias and tail distance bias">
##INFO=<ID=INDEL,Number=0,Type=Flag,Description="Indicates that the variant is an INDEL.">
##INFO=<ID=PC2,Number=2,Type=Integer,Description="Phred probability of the nonRef allele frequency in group1 samples being larger (,smaller) than in group2.">
##INFO=<ID=PCHI2,Number=1,Type=Float,Description="Posterior weighted chi^2 P-value for testing the association between group1 and group2 samples.">
##INFO=<ID=QCHI2,Number=1,Type=Integer,Description="Phred scaled PCHI2.">
##INFO=<ID=PR,Number=1,Type=Integer,Description="# permutations yielding a smaller PCHI2.">
##FORMAT=<ID=GT,Number=1,Type=String,Description="Genotype">
##FORMAT=<ID=GQ,Number=1,Type=Integer,Description="Genotype Quality">
##FORMAT=<ID=GL,Number=3,Type=Float,Description="Likelihoods for RR,RA,AA genotypes (R=ref,A=alt)">
##FORMAT=<ID=DP,Number=1,Type=Integer,Description="# high-quality bases">
##FORMAT=<ID=SP,Number=1,Type=Integer,Description="Phred-scaled strand bias P-value">
##FORMAT=<ID=PL,Number=-1,Type=Integer,Description="List of Phred-scaled genotype likelihoods, number of values is (#ALT+1)*(#ALT+2)/2">
#CHROM POS ID REF ALT QUAL FILTER INFO FORMAT Galaxy_rmdup.bam
13 19240775 . G A 145 . DP=42;AF1=1;CI95=1,1;DP4=0,0,5,13;MQ=20;FQ=-81 GT:PL:GQ 1/1:178,54,0:99
13 19254916 . T C 75 . DP=54;AF1=1;CI95=1,1;DP4=0,0,41,2;MQ=20;FQ=-156 GT:PL:GQ 1/1:108,129,0:99
13 19393304 . C G 109 . DP=40;AF1=0.5001;CI95=0.5,0.5;DP4=3,6,17,13;MQ=20;FQ=8.65;PV4=0.27,1e-22,1,0.2 GT:PL:GQ 0/1:139,0,35:38
13 19411958 . T C 136 . DP=55;AF1=1;CI95=1,1;DP4=0,0,40,11;MQ=20;FQ=-181 GT:PL:GQ 1/1:169,154,0:99
13 19419847 . C A 6.99 . DP=36;AF1=0.5015;CI95=0.5,0.5;DP4=2,0,3,0;MQ=20;FQ=-6.31;PV4=1,0.00044,1,1 GT:PL:GQ 0/1:36,0,22:26
13 19419969 . T C 18.1 . DP=66;AF1=0.5;CI95=0.5,0.5;DP4=0,36,0,28;MQ=20;FQ=19.9;PV4=1,8.3e-31,1,0.19 GT:PL:GQ 0/1:48,0,53:50
13 19429058 . C T 53 . DP=57;AF1=1;CI95=1,1;DP4=0,0,0,53;MQ=20;FQ=-187 GT:PL:GQ 1/1:86,160,0:99
13 19444828 . C T 222 . DP=211;AF1=1;CI95=1,1;DP4=0,0,78,115;MQ=20;FQ=-282 GT:PL:GQ 1/1:255,255,0:99
13 19467950 . A G 140 . DP=41;AF1=1;CI95=1,1;DP4=0,0,29,10;MQ=20;FQ=-144 GT:PL:GQ 1/1:173,117,0:99
13 19529661 . C T 70 . DP=32;AF1=1;CI95=1,1;DP4=0,0,24,1;MQ=20;FQ=-102 GT:PL:GQ 1/1:103,75,0:99
13 19536083 . G A 75 . DP=32;AF1=1;CI95=1,1;DP4=0,0,13,1;MQ=20;FQ=-69 GT:PL:GQ 1/1:108,42,0:81
13 19539455 . C T 113 . DP=31;AF1=1;CI95=1,1;DP4=0,0,2,23;MQ=20;FQ=-102 GT:PL:GQ 1/1:146,75,0:99
13 19625269 . A G 74 . DP=64;AF1=1;CI95=1,1;DP4=0,0,17,1;MQ=20;FQ=-81 GT:PL:GQ 1/1:107,54,0:99
13 19700922 . T A 53 . DP=37;AF1=1;CI95=1,1;DP4=0,0,9,0;MQ=20;FQ=-54 GT:PL:GQ 1/1:86,27,0:51
13 19700978 . A G 83 . DP=67;AF1=1;CI95=1,1;DP4=0,0,59,3;MQ=20;FQ=-214 GT:PL:GQ 1/1:116,187,0:99
```